DOI:
10.1007/s00464-018-6407-6

Document Version
Peer reviewed version

Citation for published version (APA):
Guni, A., Raison, N., Challacombe, B., Khan, S., Dasgupta, P., & Ahmed, K. (2018). Development of a technical checklist for the assessment of suturing in robotic surgery. *Surgical endoscopy*, *32*(11), 4402-4407. https://doi.org/10.1007/s00464-018-6407-6

Development of a technical checklist for the assessment of suturing in robotic surgery

Authors

Ahmad Guni, BSc (Hons)[1]

Nicholas Raison, MRCS[2]

Ben Challacombe, MS, FRCS (Urol)[3]

Shamim Khan, FRCS (Urol)[2]

Prokar Dasgupta, MD, FRCS (Urol)[2]

Kamran Ahmed, PhD, FRCS (Urol)[2]

Affiliations

[1] – GKT School of Medical Education, King's College London, Guy's Campus, St. Thomas Street, London, UK

[2] – Division of Transplantation Immunology & Mucosal Biology, Faculty of Life Sciences & Medicine, Guy's Hospital, MRC Centre for Transplantation, King's College London, London, UK

[3] – Department of Urology, Guy's and St Thomas', NHS Trust, London, UK

Corresponding Author: Nicholas Raison; Division of Transplantation Immunology & Mucosal Biology, Faculty of Life Sciences & Medicine, Guy's Hospital, MRC Centre for Transplantation, King's College London, London, UK; nicholas.raison@kcl.ac.uk. Telephone: 0207 188 5906

Abstract

Background

With the increased use of simulation for surgical training, there is a need for objective forms of assessment to evaluate trainees. The Global Evaluative Assessment of Robotic Skills (GEARS) is widely used for assessing skills in robotic surgery, but there are no recognised checklist scoring systems. This study aimed to develop a checklist for suturing in robotic surgery.

Methods

A suturing checklist for needle driving and knot tying was constructed following evaluation of participants performing urethrovesical anastomoses. Key procedural steps were identified from expert videos, while assessing novice videos allowed identification of common technical errors. 22 novice and 13 expert videos were marked on needle driving, while 18 novices and 10 experts were assessed on knot tying. Validation of the finalised checklist was performed with the assessment of 39 separate novices by an expert surgeon and compared to GEARS scoring.

Results

The internal consistency of the preliminary checklist was high (Cronbach's alpha = 0.870 for needle driving items; 0.736 for knot tying items), and after removal of poorly correlating items, the final checklist contained 23 steps. Both the needle driving and knot tying categories discriminated between novices and experts, $p < 0.005$. While the GEARS score demonstrated construct validity for needle driving, it could not significantly differentiate between novices and experts for knot tying, $p = 0.286$. The needle driving category significantly correlated with the corresponding GEARS scores ($r_s = 0.613$, $p < 0.005$), but the correlation for knot tying was

insignificant (r_s - 0.296, p = 0.127). The pilot data indicates the checklist significantly correlated with the GEARS score ($p < 0.005$).

Conclusion

This study reports the development of a valid assessment tool for suturing in robotic surgery. Given that checklists are simple to use, there is significant scope for this checklist to be used in surgical training.

Keywords: Robotic Surgery; Suturing; Checklist; Education; Assessment

Introduction

Simulation has emerged as a key training tool designed to facilitate training outside of the operating room. The skills required for robotic surgery are unique and cannot easily be adapted from open or laparoscopic surgery without appropriate training (1). It is therefore recommended that robotic surgeons practice outside of the operating room, particularly in the initial error-prone phase of the learning curve where patient safety is most at risk (2).

Simulation training tools require objective forms of assessment in order to evaluate current expertise and track progression. These include checklists and global rating scales (3), with the Global Evaluative Assessment of Robotic Surgery (GEARS) the most widely validated and used for assessment in robotic surgery (4).

It has been proposed that checklists limit the subjective component of examination as there is a strict and unambiguous list of items that must be met, reducing the risks of variations in interpretation between examiners (5). Whilst it is suggested that checklists lack flexibility in their criteria, the subjective nature of global rating scales requires that assessment is undertaken by experienced surgeons. The feasibility of using multiple experts to assess videos of numerous trainees is uncertain (6), underlining the possible use of checklists as a practical alternative.

Owing to their ease of use, checklists have formed an important component of the assessment of surgical skills in the past two decades. Checklists have been validated for surgical procedures and tasks, including laparoscopic cholecystectomy (7), and various basic surgical skills (8). Relevant to this study, a checklist was validated for the objective assessment of laparoscopic suturing using twenty-nine items to capture the technical steps in needle driving and knot tying (9).

Checklists are simple to use and administer, do not necessarily require experts to use them, and have demonstrated reliability and validity in the literature. Despite this, unlike laparoscopic surgery, there is currently no technical checklist for suturing in robotic surgery (9). The aim of this study was to devise a checklist for suturing in robotic surgery, and establish its reliability, construct validity and concurrent validity using GEARS as the standard.

Materials and Methods

Checklist development

To construct the items in the checklist, videos of expert and novice surgeons undergoing suturing evaluation within a simulated operating room environment were reviewed. All participants performed the same task involving a robotic urethrovesical anastomosis (UVA) using a synthetic dry lab model (3d med). Novices were defined as having no laparoscopic, robotic or operative experience, while experts had performed over 50 robotic-assisted laparoscopic prostatectomy independently.

The checklist was developed in two stages. First, all potential steps of suturing relating to needle positioning, insertion into the tissue, exit out of the tissue, and knot tying were identified and recorded through a review of 13 videos of experts proficient in robotic surgery. 22 novice videos were then evaluated to identify additional key steps and frequent technical errors in needle driving, while a separate set of 18 novices were assessed in the same way for knot tying. The focus in this phase was to identify any errors that the novices commonly made, and attempt to capture this in the checklist to maximise the discriminatory power of the checklist between subjects of different skills levels. It was noted that novices commonly required multiple attempts at a single step, or made multiple errors in the same step, so sub-steps were included within certain items of the checklist to capture more than one attempt or error.

To ensure constant assessment, three final complete sutures from each attempt were marked, with the scores averaged to provide a more reliable representation of the subject's ability at needle driving. A 'general' domain was included in the needle driving category to account for overall competence, and to allow the recording of errors. In the knot tying category, the steps for the first knot are detailed and marked individually. All subsequent knots score one point for each completed step in order to accommodate for differences in the number of knots carried out depending on the technique used by the participant.

All potential key steps were compiled into a comprehensive checklist by AG after agreement with authors, using the template developed by Moorthy et al. (9). The needle driving category was used to evaluate 22 novices and 13 experts. 10 of the experts were used to evaluate the knot tying category along with a separate group of 18 novices. Assessment involved marking each step separately based on the checklist. An expert surgeon not involved in the development of the checklist acted as an independent observer and marked the participants using GEARS score. The finalised checklist underwent pilot validation. A surgeon experienced in robotic training used the checklist to assess the performances of 39 novices in two video sets. The novices had undergone robotic surgical training as part of two independent studies. Scores were compared to GEARS scores marked by an expert surgeon.

Statistical Analysis

As the needle driving and knot tying categories of the checklist were constructed from, and used to analyse, separate sets of novice videos, their analysis was carried out separately.

Cronbach's Alpha was used to determine internal consistency. Poorly correlating items were considered for deletion and discussed by the study authors. To establish construct validity, a Mann-Whitney U test was carried out to determine whether the checklist could significantly differentiate between novices and experts. Spearman's rank correlation measured the correlation between the checklist scores and GEARS.

Results

In the initial stage of constructing the checklist, 29 steps were identified, divided into 19 items related to needle driving and 10 to knot tying (Supplementary Table 1). The needle driving category contained 8 procedure-specific steps and 11 general competency items.

Reliability

Both categories demonstrated a high level of internal consistency. The Cronbach alpha of the needle driving items was 0.870, and 0.736 for the knot tying category. A Cronbach alpha above 0.70 is considered to be a satisfactory internal consistency (10).

After viewing the individual Cronbach alpha results nine steps or sub-steps in the needle driving category were identified and deemed redundant by the authors as they did not capture discriminatory elements of the suturing process (Supplementary Table 2). The same process was applied to the knot tying category. Two items were not eligible for analysis as every participant scored the same result, while removal of five other items would have led to an increased Cronbach Alpha score. It was agreed that four of these items were unnecessary, so they were removed (Supplementary Table 3). The items 'instruments positioned with correct C or reverse C loop' and 'short tail of thread is pulled completely through loop in one smooth motion' were retained in the checklist despite scoring poorly on analysis, as they were considered to be important distinct procedural steps in the knot tying process. Furthermore, the item 'No injuries to tissue in the process of knot tying' also scored poorly but was kept as it could indicate poor instrumental control, and therefore potentially discriminate between different skill levels. The completed checklist used for analysis of validity included 23 steps with a maximum score of 32 (Table 1).

Following completion of the checklist, the internal consistency was tested again. The new Cronbach alpha for the needle driving category was 0.887 and the knot tying category was 0.752, demonstrating that the internal consistency improved after removal of surplus items.

Validation

The mean scores of novices and experts as marked by both GEARS and the checklist along with the Mann-Whitney U test p value are shown in Tables 2 and 3. Both the needle driving category and the GEARS score could significantly differentiate between novices and experts, but only the knot tying checklist could differentiate between novices and experts, while the GEARS score could not (p = 0.286).

A Spearman's rank-order correlation test was carried out to measure the relationship between the checklist and the GEARS scores. The needle driving category and the accompanying GEARS scores correlated strongly (r_s - 0.616, p < 0.005), while the knot tying category and the GEARS score did not correlate significantly (r_s - 0.296, p = 0.127). Furthermore, the checklist and GEARS scores marked by the expert observer correlated significantly for both the first (r_s - 0.738, p < 0.005, n = 19) and second (r_s - 0.829, p < 0.005, n = 20) novice video sets.

Discussion

Validated checklists and global rating scores have been developed for the assessment of simulation in laparoscopic surgery (4, 11). Although a global rating system is in use for robotic surgery, there is no current checklist that can evaluate suturing in robotic surgery. This study has aimed to construct a checklist for suturing in robotic surgery.

It has been suggested that checklists decrease the interpretation component of examining, instead forcing examiners to be observers of behaviours (12). Checklists are a more unambiguous measure of performance. Provided the observer understands the items of the checklist and the significance of each item, it does not necessarily require experts to act as the observer. On the other hand, global rating scores require experienced surgeons to decide competency within a more broadly defined domain. This judgement relies on the surgeon's experience as to what constitutes competency, so a more junior surgeon or novice may not be able to assess this as accurately. Although it is desirable for expert surgeons to be observers of performance in all instances, it is not always feasible, given the time constraints, when assessing many trainees (3). Consequently, the first advantage of a checklist scoring system is that observers with different levels of experience can adequately use it provided they are appropriately informed about the steps in the procedure.

Because global rating scores depend on domain-based objectives, the GEARS score has been used to assess various surgical tasks, including robot-assisted laparoscopic prostatectomy, partial nephrectomy and continuous suturing (4, 13, 14). In contrast, checklists are based on the steps of one procedure, and are therefore more specific. If a participant is fluent in utilising the instruments but misses key steps, this may not be reflected sufficiently in the global rating scale. Checklists remove this element of subjectivity. It has long been recognised that trainees are better able to improve their proficiency with objective feedback (15). Providing a step-by-step guide to the participant on their performance accurately allows trainees to identify which specific aspect of the task to focus on. In addition, the checklist devised in this study has items at the end which cover important general competencies not covered by the previous procedural steps in the checklist, allowing a more comprehensive assessment of the participant's ability. This is a feature lacking in other checklists in the literature.

It has equally been argued that checklists lack the flexibility to recognise alternative approaches to tasks and problem solving, which are hallmarks of competent physicians (5, 12). However, a recent systematic review comparing the evidence for checklists and global rating scales in a variety of simulation models found that the pooled checklist and global rating scores correlated positively with each other, with both displaying high interrater reliability and a similar ability to discriminate between trainees of different experience levels (16). This is provided the checklist used is specific to the task undertaken, corroborating with the results reported in this study that the checklist designed can significantly differentiate between participants of differing abilities in suturing. Although the knot tying category of the checklist failed to significantly correlate with the GEARS score, this is unsurprising given that the GEARS score was not able to discriminate significantly between experts and novices. This may be because of the inherent subjectivity with global rating scores and the lack of specificity to a particular surgical task. However, when the separate set of novice videos were assessed by the expert robotic surgeon, there was a strong correlation between the checklist and GEARS score.

In a wider context, the checklist can be utilised not only for assessing the ability of trainee surgeons, but also for advances in surgical education research. The field of simulation in surgery is continually evolving, with novel technologies and strategies being developed and validated constantly to enhance the training of trainees. The checklist offers an objective scale to measure changes in a trainee's ability, allowing researchers to compare different training programs or curricula.

A number of limitations to this study need to be addressed. First, there was only observer reviewing the videos at a time. Interrater reliability can be a useful component in measuring the reliability of an assessment tool (17). However, inclusion and exclusion of checklist items required agreement among authors, allowing both expert and non-expert input into the creation of the checklist. Furthermore, an expert robotic surgeon independently reviewed two sets of suturing videos to validate the checklist. Another limitation is that the checklist has only been used in a simulated environment. Further validation will require use of the checklist during suturing in the operating room.

Conclusion

This study reports the development of a new assessment tool for suturing in robotic surgery, and demonstrates both reliability and validity in evaluating suturing skills in robotic surgery. Given that checklists are simple to use and may be used accurately by non-experts, there is significant scope for these checklists to be used in training to assess ability and track progress.

Acknowledgements

N.R and P.D acknowledge educational research support from the Vattikuti Foundation. P.D. and K.A. acknowledge educational research support from the Urology Foundation, Olympus and the Royal College of Surgeons of England. P.D. acknowledges support from the MRC

Centre for Transplantation and NIHR Biomedical Research Centre at King's College London. All other authors declare no conflict of interest.

Disclosures

Mr Ahmad Guni, Dr Nicholas Raison, Dr Ben Challacombe, Prof Shamim Khan, Prof Prokar Dasgupta and Dr Kamran Ahmed have no conflicts of interest or financial ties to disclose.

References

1. Bric J, Lumbard D, Frelich M, Gould J (2015) Current state of virtual reality simulation in robotic surgery training: a review. Surg Endosc. 30(6):2169-2178.

2. Scott DJ, Dunnington GL (2008) J Gastrointest Surg. 12(2):213-21

3. Moorthy K, Munz Y, Sarker S, Darzi A (2003) Objective assessment of technical skills in surgery. BMJ 327(7422):1032-1037.

4. Goh A, Goldfarb D, Sander J, Miles B, Dunkin B (2012) Global Evaluative Assessment of Robotic Skills: Validation of a Clinical Assessment Tool to Measure Robotic Surgical Skills. J Urol 187(1):247-252.

5. Martin J, Regehr G, Reznick RK, Macrae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84(2):273-278.

6. Torkington J, Smith S, Rees B, Darzi A (2001) Skill transfer from virtual reality to a real laparoscopic task. Surg Endosc 15(10):1076-1079.

7. Eubanks T, Clements R, Pohl D, Williams N, Schaad D, Horgan S, Pellegrini C (1999) An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 189(6):566-574.

8. Anastakis D, Regehr G, Reznick R, Cusimano M, Murnaghan J, Brown M, Hutchinson C (1999) Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 177(2):167-170.

9. Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A. (2004) Bimodal assessment of laparoscopic suturing skills: Construct and concurrent validity. Surg Endosc 18(11):1608-1612.

10. Bland J, Altman D (1997) Statistics notes: Cronbach's alpha. BMJ. 314(7080):572-572.

11. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 190(1):107-113.

12. Regehr G, MacRae H, Reznick RK, Szalay D (1998) Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 73(9):993-7.

13. Hung AJ, Bottyan T, Clifford TG, Serang S, Nakhoda ZK, Shah SH, Yokoi H, Aron M, Gill IS (2017) Structured learning for robotic surgery utilizing a proficiency score: a pilot study. World J Urol 35(1):27-34.

14. Sánchez R, Rodríguez O, Rosciano J, Vegas L, Bond V, Sanchez-Ismayel A (2016) Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS). J Robot Surg 10(3):227-231.

15. Kopta JA (1971) An approach to the evaluation of operative skills. Surgery 70: 297-303.

16. Ilgen JS, Ma IW, Hatala R, Cook DA (2015) A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ 49(2):161-173.

17. McHugh M (2012) Interrater reliability: the kappa statistic. Biochemia Medica. 76-282.

Table 1: Final checklist used for validity analysis

Needle Driving	**Step**	**Sub-step**	**1**
1	Needle loaded at ½ to ⅓ from needle driver tip	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
2	Needle inserted at 90°	± 10°	
		± 20°	
3	Points of entry	1 attempt	
		≤ 2 attempts	
4	Needle driven through in one movement		
5	Needle pulled out along its curve		
6	Stabilisation of tissue		
7	Injuries to tissue in process of needle driving	none	
		≤ 1	
		≤ 2	
8	No instrument clashes		
General			
9	Equidistant suture placement		
10	Camera view centred		
11	No suture entanglement		
12	Continuity/no hesitation		
13	Competent use of both hands		
14	Progression		
Knot Tie			
15	Instruments positioned with correct C or reverse C loop		
16	Thread wrapped around needle driver (once or twice according to technique)	1 attempt	
		≤ 2 attempts	
17	Short tail of thread is pulled completely through loop in one smooth motion		
18	For all subsequent knots, reverse of prior C loop formed		
19	For all subsequent knots, thread wrapped around needle driver (once or twice according to technique)	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
20	For all subsequent knots, short tail of thread is pulled completely through loop in one smooth motion		
21	All throws squared		
22	Needles cut from thread		
23	No injuries to tissue in process of knot tying		

Overall score _____ / 32

Table 2: Needle driving scores for the novice and expert videos evaluated

Score	**Group**	**Mean score ± SD**	**Mann-Whitney U test p value**
Needle driving category	Novice	11.80 ± 3.18	< 0.005
	Expert	17.98 ± 2.00	
GEARS	Novice	12.52 ± 3.57	< 0.005
	Expert	20.00 ± 4.24	

Table 3: Knot tying scores for the novice and expert videos evaluated

Score	**Group**	**Mean score ± SD**	**Mann-Whitney U test p value**
Knot-tying category	Novice	6.22 ± 1.96	< 0.005
	Expert	10.00 ± 1.41	
GEARS	Novice	18.28 ± 3.01	0.286
	Expert	20.40 ± 4.43	

Supplementary Table 1: Preliminary checklist

Needle Driving	**Step**	**Sub-step**	**1**
1	Needle loaded at ½ to ⅓ from needle driver tip	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
2	Needle inserted at 90°	± 10°	
		± 20°	
		± 30°	
3	Points of entry	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
4	Needle driven through in one movement		
5	Needle pulled out along its curve		
6	Stabilisation of tissue		
7	Injuries to tissue in process of needle driving	none	
		≤ 1	
		≤ 2	
8	Instrument clashes	0	
		≤ 1	
		≤ 2	
General			
9	Sutures tightened		
10	Equidistant suture placement		
11	Camera view centered		
12	Appropriate camera distance		
13	No suture entanglement		
14	Needle not dropped onto models		
15	Needle not cut/broken		
16	Needle not bent		
17	Continuity/no hesitation		
18	Competent use of both hands		
19	Progression		
Knot Tie			
20	Instruments positioned with correct C or reverse C loop		
21	Thread wrapped around needle driver (once or twice according to technique)	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
22	Short tail of thread is pulled completely through loop in one smooth motion		
23	For all subsequent knots, reverse of prior C loop formed		

24	For all subsequent knots, the thread wrapped around needle driver (once or twice according to technique)	1 attempt	
		≤ 2 attempts	
		≤ 3 attempts	
25	For all subsequent knots, the short tail of thread is pulled completely through loop in one smooth motion		
26	All throws squared		
27	Needles cut from thread		
28	No excess tension on suture (no fraying)		
29	Injuries to tissue in process of knot tying	0	
		≤ 1	
		≤ 2	

Supplementary Table 2: Analysis of internal consistency for needle driving components

Step in needle driving category	Alpha if item deleted
2 Needle inserted at 90° ± 30° (sub-step 3)	0.871*
3 Points of entry ≤ 3 (sub-step 3)	0.871*
8 Instrument clashes ≤ 1 (sub-step 2)	0.871*
8 Instrument clashes ≤ 2 (sub-step 3)	0.871*
9 Sutures tightened	0.871*
12 Appropriate camera distance	0.872*
14 Needle not dropped onto models	0.872*
15 Needle not cut/broken	0.881*
16 Needle not bent	0.871*

*****items deleted after agreement with authors

Supplementary Table 3: Analysis of internal consistency for knot tying components

Item in knot tying category	Alpha if item deleted
20 Instruments positioned with correct C or reverse C loop	0.742
21 Thread wrapped around needle driver (once or twice according to technique) ≤ 3 attempts (sub-step 3)	0.745*
22 Short tail of thread is pulled completely through loop in one smooth motion	0.764
28 No excess tension on suture (no fraying)	0.742*
29 No Injuries to tissue in process of knot tying (sub-step 1)	0.749
29 Injuries to tissue in process of knot tying ≤ 1 (sub-step 2)	Not analysed*
29 Injuries to tissue in process of knot tying ≤ 2 (sub-step 3)	Not analysed*

*Items deleted after agreement with authors